Working Papers of the Priority Programme 1859
Experience and Expectation.
Historical Foundations of Economic Behaviour
Edited by Alexander Nützenadel und Jochen Streb

No 21 (2020, September)

Foltas, Alexander / Pierdzioch, Christian

On the Efficiency of German Growth Forecasts: An Empirical Analysis Using Quantile Random Forests

Arbeitspapiere des Schwerpunktprogramms 1859 der Deutschen Forschungsgemeinschaft
„Erfahrung und Erwartung. Historische Grundlagen ökonomischen Handelns“ /
Working Papers of the German Research Foundation's Priority Programme 1859
"Experience and Expectation. Historical Foundations of Economic Behaviour"

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Published in co-operation with the documentation and publication service of the Humboldt University, Berlin (https://edoc.hu-berlin.de).

ISSN: 2510-053X

Redaktion: Alexander Nützenadel, Jochen Streb, Ingo Köhler

V.i.S.d.P.: Alexander Nützenadel, Jochen Streb

SPP 1859 "Erfahrung und Erwartung. Historische Grundlagen ökonomischen Handelns"
Sitz der Geschäftsführung:
Humboldt-Universität
Friedrichstr. 191-193, 10117 Berlin
Tel: 0049-30-2093-70615, Fax: 0049-30-2093-70644
Web: https://www.experience-expectation.de
Koordinatoren: Alexander Nützenadel, Jochen Streb
Assistent der Koordinatoren: Ingo Köhler

Recommended citation:

Foltas, Alexander / Pierdzioch, Christian (2020): *On the Efficiency of German Growth Forecasts: An Empirical Analysis Using Quantile Random Forests*. Working Papers of the Priority Programme 1859 "Experience and Expectation. Historical Foundations of Economic Behaviour" No 21 (September), Berlin

On the Efficiency of German Growth Forecasts: An Empirical Analysis Using Quantile Random Forests

Alexander Foltas[a*] and Christian Pierdzioch[a]

September 2020

Abstract

We use quantile random forests (QRF) to study the efficiency of the growth forecasts published by three leading German economic research institutes for the sample period from 1970 to 2017. To this end, we use a large array of predictors, including topics extracted by means of computational-linguistics tools from the business-cycle reports of the institutes, to model the information set of the institutes. We use this array of predictors to estimate the quantiles of the conditional distribution of the forecast errors made by the institutes, and then fit a skewed t-distribution to the estimated quantiles. We use the resulting density forecasts to compute the log probability score of the predicted forecast errors. Based on an extensive in-sample and out-of-sample analysis, we find evidence, particularly in the case of longer-term forecasts, against the null hypothesis of strongly efficient forecasts. We cannot reject weak efficiency of forecasts.

JEL classification: C53; E32; E37

Keywords: Growth forecasts; Forecast efficiency; Quantile-random forests; Density forecasts

Address:

[a] Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany

[*] Corresponding author. E-mail: foltasa@hsu-hh.de.

Funding:
This research was supported by the German Science Foundation (Project: Exploring the experience-expectation nexus in macroeconomic forecasting using computational text analysis and machine learning; Project number: 275693836).

1 Introduction

The efficiency of macroeconomic forecasts requires that information available to a forecaster when a forecast is being made do not help to explain the subsequently realized forecast error. The classic approach to test forecast efficiency is to set up a regression equation that features as dependent variable the forecast error and predictor variables that represent a forecaster's information set. Such a regression equation can then be estimated by the ordinary-least-squares technique, and standard methods can be used to test whether the estimated coefficients of the equation are not significantly different from zero (see Mincer and Zarnowitz, 1969; Holden and Peel, 1990).

We go beyond the classic approach in that we use quantile-random forests Meinshausen (2006) to re-examine the efficiency of the growth forecasts published by three leading German economic research institutes during the sample period from 1970 to 2017. For the purpose of our research, quantile-random forests have the advantage that they are a flexible data-driven modeling framework that makes it possible to proxy the research institutes information set by means of a large array of predictors. As predictors, we consider numerous macroeconomic variables often studied in earlier forecasting literature and, in addition, topics extracted by means of computational-linguistics tools from the business-cycle reports of the research institutes (Foltas, 2020). The business-cycle reports of the research institutes are likely to reflect, on the one hand, the evolution of other macroeconomic predictors. However, the business-cycle reports, on the other hand, also embed the research institutes' perception of current and future macroeconomic developments and, thereby, potentially draw a more complete picture of the research institutes' information set than standard macroeconomic predictors alone can do.

Our research contributes to recent research that uses tree-based methods to study macroeconomic forecasts for Germany. Behrens, Pierdzioch, and Risse (2018a) analyze the joint efficiency of macroeconomic forecasts by means of multivariate random forests. Behrens, Pierdzioch, and Risse (2018b) estimate random classification forests to test optimality of macroeconomic forecasts under flexible loss, and Behrens, Pierdzioch, and Risse (2019) use Bayesian trees to study

the efficiency of forecasts. They reject, using Bayesian trees, strong efficiency of forecasts and weak efficiency of longer-term forecasts. Weak forecast efficiency requires that forecast errors cannot be predicted by means of their own lagged values, while strong efficiency requires that forecast errors cannot be predicted by means of any other predictors potentially in a forecasters information set.[1]

Unlike tree-based methods considered in earlier literature on forecast efficiency, quantile-random forests have the advantage that they resemble a standard quantile-regression model (see, e.g., Koenker, 2005) insofar as the informational content of the predictors can be traced out along the quantiles of the conditional distribution of forecast errors. Like Adrian, Boyarchenko, and Giannone (2019), we use the quantiles to estimate the parameters of a skewed t-distribution. The fitted skewed t-distribution, in turn, renders it possible to produce density forecasts that we use, in an in-sample and out-of-sample analysis, to compute a sequence of the log probability scores of the predicted forecast errors. We define the research institutes forecasts to be efficient if the sequence of log probability scores is on average not different from that generated by a naive quantile-regression model that uses only quantile-specific constants to model the conditional density of the forecast errors. We use the test proposed by Amisano and Giacomini (2007) to test formally whether the difference between the sequences of probability scores is significant in a statistical sense.

We organize the remainder of this research as follows. In Section 2, we describe the quantitative methods that we use in our empirical research. In Section 3, we briefly describe our data. In Section 4, we summarize our empirical results. Based on in-sample results, we reject the null hypothesis of strongly efficient short-term and long-term forecasts. Results of an extensive out-of-sample analysis, in turn, lead us to reject strong efficiency of long-term forecasts, but not of short-term forecasts. Evidence based on both our in-sample and out-of-sample analysis against

[1]Other aspects of macroeconomic forecasts for Germany have been studied in recent research by, for example, Heilemann and Stekler (2013), who focus on the time-varying accuracy of forecasts, Kirchgässner and Müller (2006), who analyze costly forecast revisions, and Döpke and Fritsche (2006), who use panel-data methods to show that macroeconomics forecasts are unbiased and weakly efficient.

the null hypothesis of weak efficiency of forecasts is weak and mostly insignificant. In Sample 5, we offer some concluding remarks.

2 Modeling Framework

Regression trees are nonparametric models that subdivide the predictor space into nonoverlapping regions that represent a relatively homogenous outcome of the dependent variable (for a textbook introduction, see Hastie, Tibshirani, and Friedman, 2009). A regression tree consists of three main elements: an initial node (that is, a root), interior nodes, and terminal nodes (leaves).[2]

The nodes partition the predictor space, $\mathbf{X}_t$, $t = 1, ..., N$, into rectangular regions in a binary top-down way, with leaves representing a subspace of a forecast error, e_{t+h}, where h denotes the forecast horizon. The formation of the subspaces starts at the root by choosing a partitioning predictor, s, and a partitioning point, z, to form the two regions $R_1(s,z) = \{\mathbf{X}_{t,s} | \mathbf{X}_{t,s} \leq z\}$ and $R_2(s,z) = \{\mathbf{X}_{t,s} | \mathbf{X}_{t,s} > z\}$. The split is identified by solving $\min_{s,z}\{RSS_1 + RSS_2\}$, where $RSS_k = \sum_{\mathbf{X}_{t,s} \in R_k(s,z)} (e_{t+h} - \bar{e}_{t+h,k})^2$, with $\bar{e}_{t+h,k} = \text{mean}\{e_{t+h} | \mathbf{X}_{t,s} \in R_k(s,z)\}, k = 1,2$, $\mathbf{X}_{t,s} \in R_k$ denotes that the period-t realization of predictor s belongs to region R_k. The regression tree is formed by recursively applying this search-and-split approach in a hierarchical manner until some maximal node size is reached or the leaves contain a minimum number of observations, both being defined as hyperparameters by a researcher in advance.

A single regression tree has a poor forecasting performance because its hierarchical structure gives rise to a high data sensitivity. A random forest model improves forecast performance by growing a large number of independent random regression trees, whose predictions are then averaged. Each random tree that is part of the random forest is estimated on a bootstrapped sample of the data and only a random subset of the predictors is used for splitting (Breiman, 2001).

[2] Our description of regression trees is relatively compact. For a more detailed description and numerical examples, see Behrens, Pierdzioch, and Risse (2018a).

Random forests, thereby, decorrelate the predictions from individual trees, curb the influence of influential individual predictors, and thus lower the prediction variance.

As shown by Meinshausen (2006), random forests can be extended to compute the conditional distribution function of the predicted variable. Starting point is the insight that every observation receives a weight $w_t = \mathbf{1}_{\{\mathbf{X}_{t,s} \in R_k(s,z)\}} / (\#\{j : \mathbf{X}_{j,s} \in R_k(s,z)\})$ at the leaves of a regression tree, where $\mathbf{1}$ is the indicator function. The prediction of the forecast error is then $\hat{e}_{t+h} = \sum_{t=1}^{N} w_t e_{t+h}$. For a random forest, the weights are given by $w_t^B = B^{-1} \sum_{i=1}^{B} w_t$, such that $\hat{e}_{t+h} = \sum_{t=1}^{N} w_t^B e_{t+h}$, and B as the number of bootstrapped simulations. Using these weights, a quantile random forest stores all information instead of only providing the mean forecast error, such that the conditional distribution function of the forecast error can be estimated as $\hat{P}(e_{t+h} \leq e|\mathbf{X}_t) = \hat{F}(e|\mathbf{X}_t) = \sum_{t=1}^{N} w_t^B \mathbf{1}_{e_{t+h} \leq e}$. The α-quantile of the conditional distribution is given by the point where the probability that the forecast error is smaller than Q_α, given $\mathbf{X}_t$, equals α, which is estimated as $\hat{Q}_\alpha(\mathbf{X}_t) = \inf\{e : \hat{F}(e|\mathbf{X}_t) \geq \alpha\}$.

Building on recent research by Adrian, Boyarchenko, and Giannone (2019), we use the estimated quantiles for $\alpha = \{0.05, 0.25, 0.75, 0.95\}$ to estimate the parameters of a skewed t-distribution (Azzalini and Capitanio, 2003). Upon letting $\tilde{t}$ and $\tilde{T}$ denote the probability density and the cumulative distribution function of the Student t-distribution, the skewed t-distribution is given by $f(e, \mu, \sigma, v, \alpha) = \frac{2}{\sigma} \tilde{t}\left(\frac{e-\mu}{\sigma}, v\right) \tilde{T}\left(\alpha \frac{e-\mu}{\sigma} \sqrt{\frac{v+1}{v+\left(\frac{e-\mu}{\sigma}\right)^2}}, v+1\right)$, where $\mu \in \mathbb{R}$ is a location parameter, $\sigma \in \mathbb{R}^+$ is a scale parameter, $v \in \mathbb{Z}^+$ is a fatness parameter, $\alpha \in \mathbb{R}$ is a shape parameter, and where we have dropped the time subscript for notational convenience. We estimate, by means of an exactly identified system, the four parameters by minimizing the sum of squared differences between the four estimated quantiles and the corresponding quantiles of the skewed t-distribution.

Having estimated the parameters of the skewed t-distribution, we combine the resulting density forecast with the ex-post realized value of the forecast error to compute a sequence of log probability scores. In the context of our analysis, forecast efficiency requires that this sequence of log probability scores is not significantly different from the sequence of log probability scores

generated by means of a benchmark model. Our benchmark model is a naive quantile-regression model (Koenker, 2005). This model only uses quantile-specific constants to model the conditional density of the forecast errors.

Finally, we use the test proposed by Amisano and Giacomini (2007) (henceforth AG test; no weighting), which closely resembles the familiar Diebold and Mariano (1995) test, to compare formally the difference between the sequence of log probability scores implied by quantile random forests with the sequence of log probability scores implied by the benchmark model.

3 The Data

We study the annual growth forecasts of three major German economic research institutes for the sample period 1970−2017.[3] The forecast publication frequency varies across the research institutes and also over time. Most commonly available are one-year-ahead annual forecasts ($q4$-forecasts) published at the turn of the year, and six-month-ahead annual forecasts ($q2$-forecasts) published mid-year. We pool all forecasts into one sample and subtract the realized growth rate (measured using first-release data retrieved from the German statistical office) from the forecasts in order to compute the forecast errors. We adjust our data on realized growth for each institute for German reunification (for further details, see Behrens, Pierdzioch, and Risse, 2018a).

Table 1 summarizes descriptive statistics of the forecast errors. The number of forecast errors ranges between 117 and 135, depending on the forecast horizon. The $q4$-forecasts errors are negative on average, while the $q2$-forecasts have a positive mean. As one would have expected, the standard deviation (SD) and the root mean square error (RMSE) are larger for $q4$-forecasts than for $q2$-forecasts. The set of macroeconomic predictors that we use to proxy the information set of the research institutes has been studied extensively in recent research of macroeconomic

[3]The research institutes are: Deutsches Institut für Wirtschaftsforschung, Ifo Institut, and Institut für Weltwirtschaft.

Table 1: Descriptive statistics of forecast errors.

Forecasts	Mean	N	SD	RMSE
$q2$	-0.12	117	0.82	0.83
$q4$	0.05	135	1.27	1.27

N: Number of observations. SD: Standard deviation. RMSE: Root-mean-squared error.

forecasts for Germany (see, e.g., Behrens, Pierdzioch, and Risse, 2018a, 2019). The macroeconomic predictors are available at a monthly frequency, where we take into account a forecast formation lag (that is, the the research institutes use macroeconomic data for the month preceding the month in which a forecast is formed) and publication lags. The array of macroeconomic predictors includes the following variables: a short term interest rate, the term spread, the returns on the OECD share-price index for Germany, the U.S. federal funds rate, the inflow of industrial orders, the growth rate of German industrial production, the growth rate of U.S. industrial production, business tendency surveys for manufacturing (tendency and future tendency), the CPI inflation rate, the growth rate of money supply (M1), the exchange rate of the US dollar vis-à-vis the euro (before 1999, vis-à-vis the Deutsche Mark), the returns of the oil price (West Texas Intermediate), the returns of the real effective exchange rate, and the normalized OECD composite leading indicator for Germany.[4]

We supplement our macroeconomic predictors with textual predictors computed using the forecast reports of the research institutes. We use machine-learning techniques to discover semantic patterns that reflect underlying topics that got combined to form the document. The most basic topic model is the latent Dirichlet allocation (LDA) (Blei, Ng, and Jordan, 2003). The idea of LDA is that each document contains a distribution over latent topics, which contain a distribution over words. The respective topic proportions provide a low-dimensional representation of the content of each document. We combine LDA with word embedding, a method of mapping words in vector space and thus representing their meaning (Panigrahi, Simhadri, and Bhat-

[4]In order to account for data revisions, we use a backward-looking moving-average of order 12 to smooth out the effects of retrospective data revisions (CPI, M1, real effective exchange rate, industrial production, and orders).

tacharyya, 2019). The topic proportions of 24 topics are used as predictors. See Foltas (2020) for an extensive description of this combined approach and an analysis of the different topics.

4 Empirical Analysis

4.1 Methodological Issues

We use 1500, 2000, and 2500 random trees to grow random forests. As our benchmark calibration, we set the minimum node size to five, and the number of predictors randomly chosen for splitting to round(number of predictors/3), which are both default values widely used in the machine-learning literature. We use the R Core Team (2020) programming environment for statistical computing to undertake our empirical analysis, where we use the add-on package "grf" (Tibshirani, Wager, and Athey, 2020) to estimate the quantile random forests and the BFGS algorithm implemented in the "optimx" package (Nash, Varadhan, and Grothendieck, 2020) to estimate the parameters of the skewd t-distribution. For estimation of the benchmark quantile-regression model, we use the "quantreg" package (Koenker, 2020, we use the "fn" algorithm).

We assess the efficiency of forecasts in the context of an in-sample and an out-of-sample analysis. For the in-sample analysis, we use the full sample of data available for our empirical analysis. We use out-of-bag data to compute forecasts, that is, the bootstrapped data not used for growing a random tree. For the out-of-sample analysis, we use a recursive and a rolling estimation approach. This approach requires that, after having started the estimations using data for some initial period, we reestimate the quantile-random forest whenever a new forecast error becomes available, where the length of the estimation window either expands (recursive) or is held fixed (rolling). For the out-of-sample analysis, we compute forecasts by plugging into an estimated random forest the new realizations of the predictors that become available when the research institutes publish a new growth forecast.

Table 2: In-sample results

Forecasts Trees	$q2$ AG-weak	$q2$ pval	$q4$ AG-weak	$q4$ pval	$q2$ AG-strong	$q2$ pval	$q4$ AG-strong	$q4$ pval
1500	-1.596	0.945	-1.065	0.857	2.888	0.002	4.797	0.000
2000	-1.926	0.973	-0.610	0.720	2.466	0.007	4.844	0.000
2500	-1.961	0.975	-0.854	0.803	2.587	0.005	4.702	0.000

AG test: Amisano-Giacomini test. A positive test statistic indicates that quantile-random forests perform better than the benchmark model. pval: one-sided p-value.

4.2 In-Sample Results

Table 2 summarizes our in-sample results, where we report for the $q2$-forecasts and the $q4$-forecasts both the value of the AG test statistic and a one-sided p-value. A positive value of the test statistic indicates that random forests perform better in terms of their implied density forecasts than the benchmark model. We compare the benchmark model, which implies that neither the lagged forecast error nor conventional macroeconomic predictors nor the LDA-based topics have predictive value for the conditional density of the forecast errors, with two versions of quantile-random forests. The first version is a very simple and stylized quantile-random forest estimated using only the lagged forecast error. We use this version to assess whether the research institutes publish weakly efficient growth forecasts. The second version features the lagged forecast error and the entire array of macroeconomic predictors and topics. It is this second version that is a natural candidate for estimation of a quantile-random forest, given the large number of predictors and the he limited number of forecasts we can use for our empirical analysis. We use this version to study whether the growth forecasts published by the research institutes are strongly efficient.

The null hypothesis is that forecasts are (weakly or strongly) efficient, while the alternative hypothesis is that forecasts are not (weakly or strongly) efficient. The main message to take home from Table 2 is that the null hypothesis of strongly efficient forecasts can be rejected at conventional levels of significance at both forecast horizon. The results of the tests for weak forecast

Table 3: In-sample results (without financial crisis)

Forecasts Trees	$q2$ AG-weak	$q2$ pval	$q4$ AG-weak	$q4$ pval	$q2$ AG-strong	$q2$ pval	$q4$ AG-strong	$q4$ pval
1500	-1.465	0.929	-1.518	0.936	2.200	0.014	4.290	0.000
2000	-1.827	0.966	-0.944	0.827	1.847	0.032	4.310	0.000
2500	-1.854	0.968	-1.194	0.884	2.106	0.018	4.247	0.000

AG test: Amisano-Giacomini test. A positive test statistic indicates that quantile-random forests perform better than the benchmark model. pval: one-sided p-value.

Table 4: In-sample results (bagging)

Forecasts Trees	$q2$ AG-strong	$q2$ pval	$q4$ AG-strong	$q4$ pval
1500	2.217	0.027	4.232	0.000
2000	2.896	0.004	3.910	0.000
2500	2.631	0.009	4.561	0.000

AG test: Amisano-Giacomini test. A positive test statistic indicates that quantile-random forests perform better than the benchmark model. pval: one-sided p-value.

efficiency, in contrast, are all insignificant and, in fact, the negative values of the AG test statistic indicate that the naive quantile-regression model yields the superior log probability scores.[5]

As a robustness check, we report in Table 3 the results we obtain when deleting the probability scores for the years 2008/2009, the years of the Great financial crisis, from our analysis.[6] This robustness check is motivated by the observation that the research institutes made relatively large forecast errors during the financial crisis. The results of this robustness check corroborate the baseline results we report in Table 2, that is, we (do not) reject (weak) strong forecast efficiency.

As another robustness check, we consider a version of random forests that uses all predictors to

[5]We also tested the weak efficiency of the forecasts by comparing the log probability score of a quantile-regression model that features only a constant with the log probability score of a quantile-regression model that features the lagged forecast error as a predictor. The findings from this comparison (not reported, but available upon request) corroborate the results reported in Table 2

[6]Recent empirical findings reported by (Döpke, Fritsche, and Müller, 2019) indicate that forecaster's behavior has changed following the financial crisis.

Table 5: Out-of-sample results

Panel A: Recursive estimation window

Forecasts	$q2$	$q2$	$q4$	$q4$	$q2$	$q2$	$q4$	$q4$
Trees	AG-weak	pval	AG-weak	pval	AG-strong	pval	AG-strong	pval
1500	-1.681	0.954	0.892	0.186	0.304	0.381	3.619	0.000
2000	-1.732	0.958	1.090	0.138	0.261	0.397	3.412	0.000
2500	-1.678	0.953	1.002	0.158	0.424	0.336	2.877	0.002

Panel B: Rolling estimation window

Forecasts	$q2$	$q2$	$q4$	$q4$	$q2$	$q2$	$q4$	$q4$
Trees	AG-weak	pval	AG-weak	pval	AG-strong	pval	AG-strong	pval
1500	-0.344	0.635	0.988	0.161	0.411	0.340	2.817	0.002
2000	0.473	0.318	1.121	0.131	1.290	0.099	2.863	0.002
2500	0.365	0.357	1.156	0.124	0.321	0.374	3.012	0.001

AG test: Amisano-Giacomini test. A positive test statistic indicates that quantile-random forests perform better than the benchmark model. pval: one-sided p-value.

grow trees. This version is also known in the machine-learning literature as "bagging" (Breiman, 1996). Table 4 summarizes the results. Bagging makes sense only when we study the version of quantile-random forests that use the full array of predictors as candidates for splitting. In other words, we focus on the case of strong forecast efficiency. We summarize the results in Table 4. The results lend further support to the view that the growth forecasts of the research institutes do not pass the test of strong efficiency.

4.3 Out-of-Sample Results

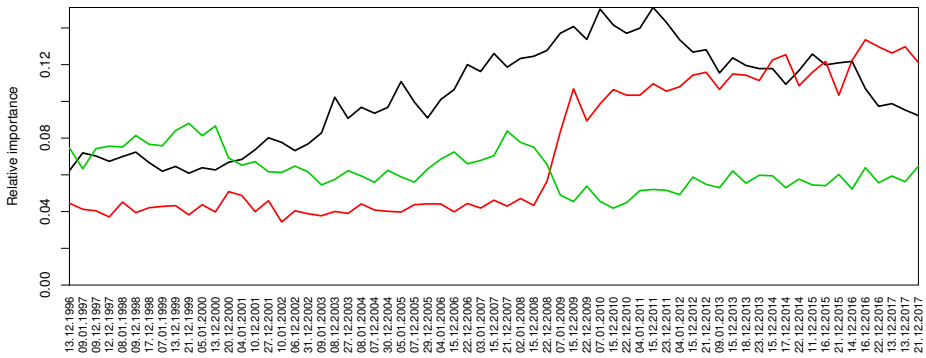
Table 5 summarizes the results of the out-of-sample analysis for both the recursive and the rolling estimation window.[7] We use a rolling estimation window of length 70 forecasts, which equals roughly half the sample size. Correspondingly, we use 70 forecasts to initialize the recursive

[7]Deleting the years of the Great financial crisis from the analysis leaves the results of the AG tests qualitatively unaffected. Results are not reported, but are available upon request.

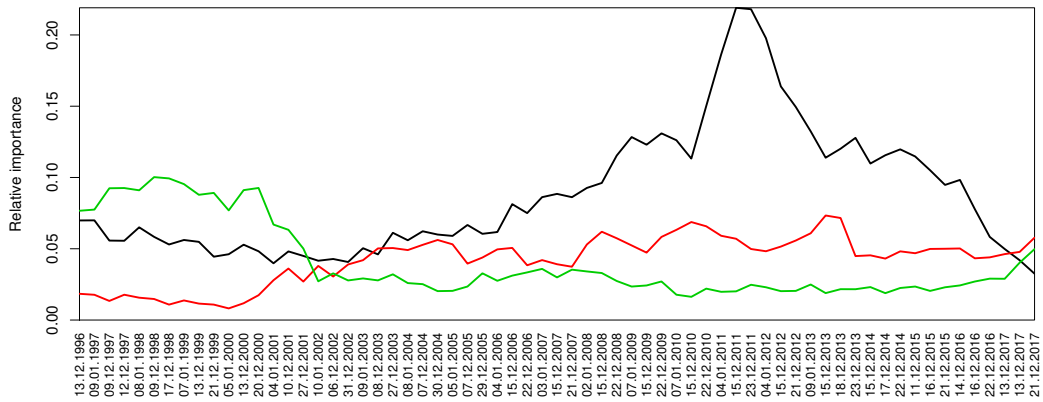
estimations. Two results stand out. First, we cannot reject weak forecast efficiency, which is in line with the in-sample results. Second, we reject strong forecast efficiency for the $q4$-forecasts but, in contrast to the in-sample analysis, not for the $q2$-forecasts. Figure 1 plots the relative

Figure 1: Variable importance

Panel A: Recursive estimation window



Panel B: Rolling estimation window

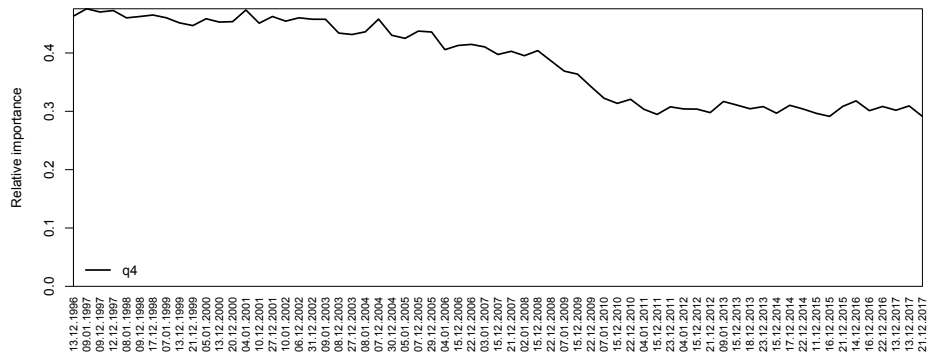


Relative importance is reported for $q4$-forecasts. Panel A: Black line − stock market returns. Red line − Business-climate expectations. Green line − returns of the real effective exchange rate. Panel B: Black line − stock market returns. Red line − growth rate of M1. Green line − returns of the real effective exchange rate. Number of trees: 2000.

importance (in percent) of the top three predictors over time for both the recursive and the rolling estimation window. Relative importance measures how often a predictor is used for splitting. We

focus on the $q4$-forecasts because we reject strong efficiency in case of these forecasts. The first result is that stock-market returns are a top predictor. The relative importance of stock-market returns increased until the financial crisis. In other words, the research institutes did not fully account for stock-market developments when forming their longer-term growth forecasts. This changed after the financial crisis, as witnessed by the decreasing relative importance of stock-market returns. The returns of the real effective exchange rate are another top predictor. The relative importance of this predictor shows a trend decline from around 8-10% to around only 4-5%, possibly reflecting that the research institutes learned over time how movements of the real effective exchange rate affected the highly export-oriented German economy. Two other noticeable predictors are business-climate expectations (recursive) and the growth rate of M1 (rolling). The relative importance of the former increased after thew Great financial crisis in case of the recursive estimation window. The relative importance of the growth rate of M1, in contrast, hovers around only 5% in case of the rolling estimation window.

Figure 2: Relative importance of the topics



Relative importance is aggregated across topics for the case of a recursive estimation window. Results are for $q4$-forecasts. Number of trees: 2000.

It is worth noting that the LDA-topics not among the top three predictors plotted in Figure 1. A natural question, therefore, is whether our results imply that reading the business-cycle reports of the research institutes is a waste of time. In order to answer this question, we plot in Figure 2, for

the case of a recursive estimation window, the relative importance aggregated across all topics ($q4$-forecasts; results for the $q2$-forecasts are similar and are not reported). It is evident from the figure that the topics, when taken together, are relatively often used as splitting variables. Their aggregate relative importance varies between roughly 30-50% and shows a tendency to decrease over time. Notwithstanding, it is fair to conclude that, while individual topics are not among the top predictors in terms of their relative importance, aggregating the relative importance of the diverse topics indicates that the business-cycle reports published by the research institutes do contain information useful for modeling the forecast errors.

4.4 Some Model Diagnostics

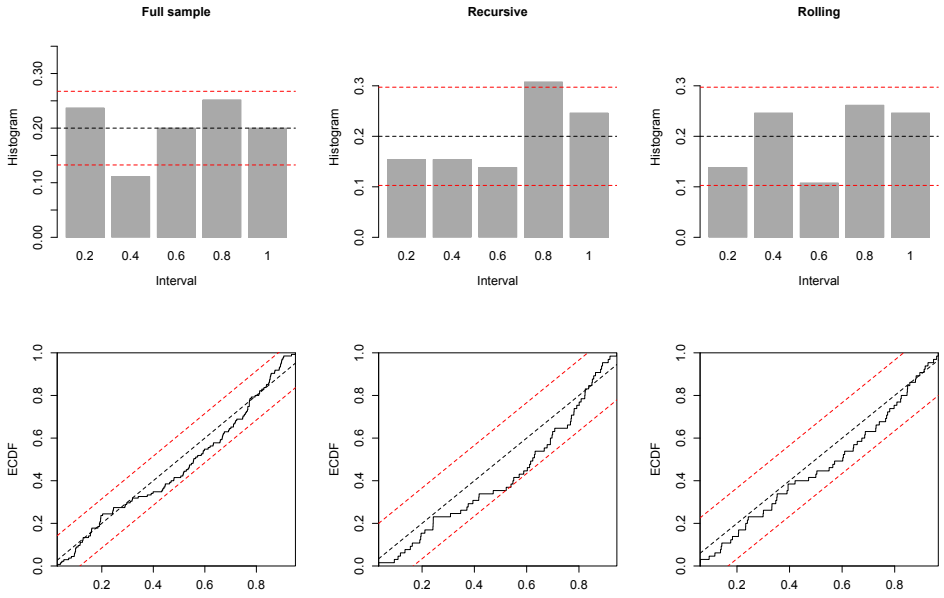
We use the probability integral transform (PIT; that is, the cumulative density function as evaluated at the actual forecast error) to assess the density forecasts. Diebold, Gunther, and Tay (1998) show that correctly specified density forecasts imply that the PIT is identically uniformly distributed on the unit interval.[8] We use the histogram of the PIT (upper row of Figure 3), the test for a correct specification of conditional predictive density by Rossi and Sekhposyan (2019) (lower row of Figure 3), and the Anderson–Darling test (Anderson and Darling, 1954) to test for uniformity of the PIT, and the stability test proposed by Andrews (1993) to test the identical-distribution property (Table 6).[9]

We focus on the $q4$-forecasts because we find strong evidence against strong efficiency of these forecasts, where we report results for the in-sample (that is, out-of-bag) and out-of-sample density forecasts. The histograms of the PITs indicate that the deviations from a uniform distribution are hardly significant. Similarly, the test for a correct specification of the conditional predictive

[8]The overlapping nature of the research institutes growth forecasts implies that, in our empirical analysis, the PIT is not independently distributed. As a result, application of the Ljung-Box test statistic yields significant results (results are not reported but available upon request).

[9]Rossi and Sekhposyan (2014) describe in detail tests useful for evaluating predictive densities.

Figure 3: Properties of the PIT



Upper row: Histograms of the PIT for $q4$-forecasts along with critical values (thin red dashed lined). Dashed red lines are the 2.5th and 97.5th percentiles bands. Lower row: Empirical cumulative distribution functions (ECDF) of the PIT (thick solid line) along with the ECDF of a uniform distribution (45-degree; thin black dashed line) line, and the 5% critical lines (red dashed lines) computed using the critical values tabulated by Rossi and Sekhposyan (2019). Number of trees: 2000.

Table 6: Model diagnostics

Test statistic	In sample	Recursive	Rolling
Anderson-Darling test	0.213	0.721	0.047
Andrews test	0.398	0.160	0.114

Model diagnostics (p-values) are reported for $q4$-forecasts. Number of trees: 2000. Anderson-Darling test: Uniformity. Andrews test: Stability.

densities yields insignificant results. The Anderson-Darling test yields insignificant results except in the case of the rolling estimation window. Similarly, the Andrews test for stability is not significant.

5 Concluding Remarks

We have used a large array of macroeconomic and textual predictors to test the weak-form and strong-form efficiency of the growth forecasts of three leading German research institutes. To this end, we have used quantile-random forests to compute density forecasts of the forecast errors. Our in-sample results show that we can reject the null hypothesis of strongly efficient short-term and longer-term growth forecasts. As for the out-of-sample results, we can reject strong efficiency of longer-term growth forecasts. We cannot reject the hypothesis of weak efficiency of growth forecasts.

On the methodological front, we have shown that quantile random forests are a useful technique for modeling a forecaster's information set with a large array of predictors even when the number of forecasts available for an in-depth empirical analysis is limited. Moreover, we have shown how the estimated quantile random forests can be used to produce density forecasts that, in turn, render it possible to analyze the efficiency of macroeconomic forecasts from a new perspective. To the best of our knowledge, this perspective has not been considered in earlier research on the efficiency of macroeconomic forecasts.

In future research, it is interesting to examine whether the efficiency of macroeconomics forecasts is related to macroeconomic risks. In this regard, one could build on recent research by Adams, Adrian, Boyarchenko, and Giannone (2020) who use quantile regressions and the skewed t-distribution to model the conditional distribution of forecast errors as implied by the median forecast of the Survey of Professional Forecasters as a function of a small number of conditioning variables (notably an index of financial conditions). They then derive from the estimated distributions metrics of downside and upside risks of key macroeconomic variables. It is interesting

to explore whether such downside and upside risks have explanatory value for the potentially time-varying efficiency of macroeconomic forecasts.

References

Adams, P. / *Adrian,* T. / *Boyarchenko,* N. / *Giannone,* D. (2020): "Forecasting Macroeconomic Risks", Federal Reserve Bank of New York. Staff Report (914).

Adrian, T. / *Boyarchenko,* N. / *Giannone,* D. (2019): "Vulnerable Growth", American Economic Review 23(4), 1263–1291.

Amisano, G. / *Giacomini,* R. (2007): "Comparing Density Forecasts via Weighted Likelihood Ratio Tests", Journal of Business and Economic Statistics 25(2), 177–190.

Anderson, T. W. / *Darling,* D. A. (1954): "A Test of Goodnes of Fit", Journal of the American Statistical Association 49(268), 765–769.

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point", Econometrica 61(4), 831–856.

Azzalini, A. / *Capitanio,* A. (2003): "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution", Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65(2), 367–389.

Behrens, C. / *Pierdzioch,* C. / *Risse,* M. (2018a): "A Test of the Joint Efficiency of Macroeconomic Forecasts Using Multivariate Random Forests", Journal of Forecasting 37(5), 560–572.

– (2018b): "Testing the Optimality of Inflation Forecasts Under Flexible Loss with Random Forests", Economic Modelling 72, 270–277.

– (2019): "Do German economic research institutes publish efficient growth and inflation forecasts? A Bayesian analysis", Journal of Applied Statistics 47(4), 698–723.

Blei, D. M. / *Ng,* A. Y. / *Jordan,* M. I. (2003): "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, 993–1022.

Breiman, L. (1996): "Bagging predictors", Machine Learning 24(2), 123–140.

– (2001): “Random forests”, Machine Learning 1(45), 5–32.

Diebold, F. / *Gunther,* T. / *Tay,* A. (1998): “Evaluating Density Forecasts, with Applications to Financial Risk Management”, International Economic Review 39(4), 863–883.

Diebold, F. X. / *Mariano,* R. S. (1995): “Comparing predictive accuracy”, Journal of Business and Economic Statistics 13(3), 253–263.

Döpke, J. / *Fritsche,* U. (2006): “Growth and inflation forecasts for Germany - A panel-based assessment of accuracy and efficiency”, Empirical Economics 31(3), 777–798.

Döpke, J. / *Fritsche,* U. / *Müller,* K. (2019): “Has macroeconomic forecasting changed after the Great Recession? Panel-based evidence on forecast accuracy and forecaster behavior from Germany”, Journal of Macroeconomics 62, 103–135.

Foltas, A. (2020): “Testing Investment Forecast Efficiency with Textual Data”, Working Papers of the Priority Programme 1859 "Experience and Expectation. Historical Foundations of Economic Behaviour" (19), `https://edoc.hu-berlin.de/handle/18452/22538`.

Hastie, T. / *Tibshirani,* R. / *Friedman,* J. (2009): “The elements of statistical learning. Data mining, inference, and prediction”, Springer: New York, 2nd edition.

Heilemann, U. / *Stekler,* H. O. (2013): “Has the accuracy of German macroeconomic forecasts improved?”, German Economic Review 14(2), 235–253.

Holden, K. / *Peel,* D. A. (1990): “On testing for unbiasedness and efficiency of forecasts”, The Manchester School 58(2), 120–127.

Kirchgässner, G. / *Müller,* U. K. (2006): “Are forecasters reluctant to revise their predictions? Some German evidence”, Journal of Forecasting 25(6), 401–423.

Koenker, R. (2005): “Quantile regression”, Cambridge University Press: Cambridge, UK.

– (2020): quantreg: Quantile Regression. R package version 5.55, `https://CRAN.R-project.org/package=quantreg`.

Meinshausen, N. (2006): "Quantile regression forests", Journal of Machine Learning Research 7, 983–999.

Mincer, J. A. / *Zarnowitz,* V. (1969): "The evaluation of economic forecasts", in: Economic forecasts and expectations: Analysis of forecasting behavior and performance, ed. by Jacob A. Mincer, 3–46, National Bureau of Economic Research, New York.

Nash, J. C. / *Varadhan,* R. / *Grothendieck,* G. (2020): optimx: Expanded Replacement and Extension of the 'optim' Function. Version 2020-4.2, `https://cran.r-project.org/web/packages/optimx/optimx.pdf`.

Panigrahi, A. / *Simhadri,* H. V. / *Bhattacharyya,* C. (2019): "Word2Sense: Sparse Interpretable Word Embeddings", Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 5692–5705.

R Core Team (2020): R: A language and environment for statistical computing, R version 3.3.3, `https://www.R-project.org/`.

Rossi, B. / *Sekhposyan,* T. (2014): "Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set", International Journal of Forecasting 30(3), 662–682.

– (2019): "Alternative tests for correct specification of conditional predictive densities", Journal of Econometrics 208(2), 638–657.

Tibshirani, J. / *Wager,* S. / *Athey,* S. (2020): grf: Generalized Random Forests. R package version 1.1.0, `https://cran.r-project.org/web/packages/grf/grf.pdf`.